

***Thema* version 1.0**

Basic User Instructions

***Thema* International Steering Committee**
October 2013

Contents

1. Introduction
2. Composition of *Thema* scheme
3. Allocation of *Thema* Subject Categories and Qualifiers
 - a. General rules of *Thema* classification
 - b. Section summary – Subject Categories
 - c. Section summary – Qualifiers
4. Worked examples
5. Communication of *Thema* Subject Categories and Qualifiers
 - a. In ONIX
 - b. In formats other than ONIX
6. Resolution of National Extensions

1. Introduction

The *Thema* scheme is intended as a subject classification scheme for book content, for use with digital and physical products, for all sectors of the global book and e-book trade, and for both online and bricks and mortar retail sectors.

The first full release of the scheme, *Thema* version 1.0, was launched in October 2013, after the Frankfurt Book Fair, and is available (initially as an Excel spreadsheet) from the EDItEUR website at <http://www.editeur.org/151/Thema/>

An interactive multi-lingual search tool is also available at <http://editeur.dyndns.org/thema/>

These Basic User Instructions are not intended to provide comprehensive guidelines, but simply to outline the fundamental principles of the application of *Thema* classification values. More extensive notes, guidelines, tools and worked examples may follow.

2. Composition of the *Thema* scheme

The *Thema* subject classification scheme comprises the following elements:

- Subject Categories, arranged over 20 sections as listed in Part 3b below, and
- Qualifiers, in six lists covering various facets, as listed in Part 3b below.

Of these elements, only a Subject Category is mandatory for the classification of a book or e-book; optional Qualifiers may be used in conjunction with Subject Categories to supplement and qualify or refine the meaning of the applied Subject Categories as required.

A book can be classified using a single Subject Category, or multiple Subject Categories, and may have none, one or many Qualifiers added.

Each Subject Category or Qualifier consists of a code and a heading, and may also include translations of the heading in other languages, heading synonyms, and associated usage notes. Codes and headings are arranged in a hierarchical pattern to simplify classification and manual searching for topics.

Subject Category codes are of variable length and always begin with a letter, *eg*

A	The Arts
FF	Crime & mystery fiction
NH	History
UGB	Web graphics & design
YBC	Children's picture books
QRRF	Zoroastrianism
LWKF	Shariah law: family relations
DNBB1	Autobiography: business & industry
MFKC3	Embryology

Qualifier codes are of variable length and always begin with an integer, *eg*

1HFGU	Uganda
2ACSC	Icelandic
3MD	16th century, c 1500 to c 1599
4GH	For International GCSE (IGCSE)
5AG	Interest age: from c 6 years
6BA	Baroque

In product metadata records and message formats such as ONIX, only the code is required to be communicated, which can then be translated by recipients as appropriate. This is important because typically a user is working with a particular language translation of the *Thema* scheme specific to a language group or country, and while the codes are universal, the heading can be presented in the user's preferred language.

The set of Subject Categories is universally applicable and common to all language translations of the *Thema* scheme. The Qualifiers contain 'shared values' which are similarly common to all translations, and may also contain 'National Extensions' which typically relate to very specific topics identified as important in certain countries, but presumed not to be of universal interest or applicability. The rules for applying National Extension Qualifiers are exactly the same as for shared value Qualifiers.

National Extension Qualifier codes are of variable length, but always have a stem which is a shared value. The stem is followed by the ISO Country Code (or an equivalent), hyphen-delimited, and then possibly by further letters: *eg*

1DNS-SE-BH	Västergötland	
3MPBGJ-ES-B	España: Guerra civil (1936–1939)	
4Z-US-	For US educational curricula	(note the trailing hyphen)
4Z-GB-SD	For Scottish Curriculum National 5	

However, not every user of *Thema* needs to ‘understand’ every National Extension: they may decide that some or all National Extensions are too detailed and choose to ignore them. The procedures for handling the receipt of ‘unknown’ National Extensions are explained in Part 6 below.

3. Allocation of *Thema* Subject Categories and Qualifiers

a. General rules of *Thema* classification

The basic rules for assigning *Thema* Subject Categories and Qualifiers, put most simply, are as follows:

Rule 1: Ensure the first Category entered is the primary subject

Taking advantage of all available information about the subject and scope of the work, ensure the first Category indicates the single most pertinent and important subject of the work.

Rule 2: Classify titles as precisely as applicable or as broadly as required

If a book is solely concerned with a specific topic defined by a single Category, use that precise code only; if it is broad-ranging in its scope within a single subject area, use a single broad Category rather than listing all the more specific Categories listed beneath it (see Note A below). Not every title has to be assigned the most detailed *possible* Category, only the most detailed that is applicable.

Rule 3: Assign as many Categories as are required

If the book covers several subject areas, assign relevant Categories from across the *Thema* scheme to indicate this, as many as are required (see Note B below). Assign only Categories for subjects that the book is substantially about as a whole, not for subjects that are peripheral.

Rule 4: Observe any Scope Notes and instructions linked to Categories.

In the hierarchical listing, many Categories (and some Qualifiers) carry specific Scope Notes and See Also references. Sometimes these are listed under a broad Category (eg at level 1 or 2 in the hierarchy) but they apply to all Categories listed hierarchically under the broad Category. The Scope Notes give important and useful information about what it is covered by a heading, and how it should be used in conjunction with other *Thema* values to make full use of the scheme and maintain consistency. The See Also references point the user to other related *Thema* values that may be appropriate to be assigned instead of, or as well as, the value under consideration.

Rule 5: Add Qualifiers whenever appropriate.

The Qualifiers are a key feature of the *Thema* scheme structure and provide great flexibility, enabling the expression of a vast range of subjects with a comparatively small number of

values. Note that the Qualifiers are associated with the book as a whole and not linked directly with a specific Category, although the link between Category and Qualifier is often implicit (*eg* Language learning material and Language Qualifiers, travel guides and Geographical Qualifiers *etc*). As with Categories, assign the most specific Qualifier values that are applicable (including National Extensions where available), as many as are required (see Notes A and B below)

Rule 6: Consider Context

Always view and consider specific headings in the context of the scheme as a whole. Sometimes a heading may seem appropriate when viewed alone, but its location in the scheme may indicate connotations that are not appropriate. This principle is particularly pertinent to selecting between similar headings that indicate different treatments of related subject matter (*eg* works for specialist or general readers, adults or children *etc*). See the Section Summary below.

Notes

Note A: Because of the hierarchical nature of the codes, it is not necessary to use ‘high-level’ codes in conjunction with lower level codes denoting specific subjects. With the specific Category QDHR3 ‘Pragmatism’ for example, it is not usually appropriate to add all or any of the preceding codes QDHR, QDH, QD or indeed Q. The fact that the subject is an aspect of QD ‘Philosophy’ is implicit in the code. Similarly, if assigning Geographical Qualifier 1HFGA ‘Ethiopia’, it is not appropriate to assign also 1H to indicate that country is in Africa, and so on.

On the other hand, if the book is about modern philosophy as a whole, with a chapter on Pragmatism, another on Phenomenology and a third on Positivism, *etc*, the correct code to apply would be the broader QDHR, not QDHR3 *and* QDHR5 *and* QDHR7.

Note B: There is no defined upper limit of the number of Subject Category values or Qualifier values that may be assigned to an individual book product record, but multiple values must be allowed for to enable the scheme to function effectively. It is suggested that a maximum of *five* of each type would more than sufficiently cover most circumstances.

b. Section summary – Subject Categories

Codes beginning A	THE ARTS
Treatment	Specialist and some General Adult; prefer W* codes for practical and hobbyist approaches
Qualifiers	Particularly STYLE and others as appropriate
Special Features	

Codes beginning C	LANGUAGE & LINGUISTICS
Treatment	Specialist and General Adult
Qualifiers	Particularly LANGUAGE and others as appropriate
Special Features	

Codes beginning D	BIOGRAPHY, LITERATURE & LITERARY STUDIES
Treatment	Specialist and General Adult
Qualifiers	Particularly LANGUAGE & STYLE, and others as appropriate
Special Features	For Literary Studies, use LANGUAGE Qualifiers to indicate the language of the literature in question; use GEOGRAPHICAL Qualifiers only to indicate national context where significant, <i>eg</i> literature in English of Australia

Codes beginning F	FICTION & RELATED ITEMS
Treatment	General Adult
Qualifiers	GEOGRAPHICAL and/or TIME PERIOD Qualifiers where these aspects are particularly pertinent & significant; SPECIAL INTEREST Qualifiers as appropriate
Special Features	FY* codes must be used with (after) F* codes

Codes beginning G	REFERENCE, INFORMATION & INTERDISCIPLINARY SUBJECTS
Treatment	Specialist and General Adult
Qualifiers	As appropriate
Special Features	GBC must be used with (after) other code denoting subject

Codes beginning J	SOCIETY & SOCIAL SCIENCES
Treatment	Specialist and some General Adult; see also V* codes for practical approaches to issues
Qualifiers	As appropriate
Special Features	JN* codes may be used with YP* codes without the need for educational-level audience codes

Codes beginning K	ECONOMICS, FINANCE, BUSINESS & MANAGEMENT
Treatment	Specialist and some General Adult; see also V* codes for practical approaches to issues
Qualifiers	As appropriate
Special Features	

Codes beginning L	LAW
Treatment	Specialist only; see V* codes for General Adult items
Qualifiers	GEOGRAPHICAL must be assigned to LN* codes
Special Features	

Codes beginning M	MEDICINE & NURSING
Treatment	Specialist only; see V* codes for General Adult items
Qualifiers	Not generally required
Special Features	

Codes beginning N	HISTORY
Treatment	Specialist and General Adult
Qualifiers	Particularly TIME PERIOD & GEOGRAPHICAL, and others as appropriate
Special Features	

Codes beginning P	MATHEMATICS & SCIENCE
Treatment	Specialist only (except see PDZ for popular treatments); see W* codes for General Adult items
Qualifiers	Not generally required
Special Features	

Codes beginning Q	PHILOSOPHY & RELIGION
Treatment	Specialist and some General Adult; prefer VX* codes for Mind, Body, Spirit approaches
Qualifiers	As appropriate
Special Features	QRV* codes should be used with (after) other QR* codes

Codes beginning R	EARTH SCIENCES, GEOGRAPHY, ENVIRONMENT, PLANNING
Treatment	Specialist and some General Adult; prefer V* and W* codes for General Adult items
Qualifiers	Particularly GEOGRAPHICAL and others as appropriate
Special Features	

Codes beginning T	TECHNOLOGY, ENGINEERING, AGRICULTURE
Treatment	Specialist only; see W* codes for General Adult items
Qualifiers	Not generally required
Special Features	

Codes beginning S	SPORTS & ACTIVE OUTDOOR RECREATION
Treatment	General Adult
Qualifiers	As appropriate
Special Features	

Codes beginning U	COMPUTING & INFORMATION TECHNOLOGY
Treatment	Specialist and General Adult
Qualifiers	Not generally required
Special Features	

Codes beginning V	HEALTH, RELATIONSHIPS & PERSONAL DEVELOPMENT
Treatment	General Adult; see Codes beginnings C, J, K, L, M, Q, R, as appropriate, for Specialist works
Qualifiers	Particularly SPECIAL INTEREST and others as appropriate
Special Features	

Codes beginning W	LIFESTYLE, HOBBIES & LEISURE
Treatment	General Adult; see A*, K*, N*, P*, R*, T* codes, as appropriate, for Specialist works
Qualifiers	Particularly SPECIAL INTEREST and others as appropriate; GEOGRAPHICAL for WT* Travel
Special Features	

Codes beginning X	GRAPHIC NOVELS, COMIC BOOKS, CARTOONS
Treatment	General Adult, plus Teenage & Children's
Qualifiers	GEOGRAPHICAL and/or TIME PERIOD Qualifiers where these aspects are particularly pertinent & significant; INTEREST AGE & SPECIAL INTEREST Qualifiers as appropriate
Special Features	Items will typically need a "type" from XA* plus a "genre" from XQ* for full classification; note inclusion of Teenage & Children's material

Codes beginning Y	CHILDREN'S, TEENAGE & EDUCATIONAL
Treatment	All Teenage & Children's material (except see X* codes above), <i>ie</i> up to c 16 years, and minimum school leaving age for Educational material
Qualifiers	Particularly INTEREST AGE and others as appropriate; particularly EDUCATIONAL PURPOSE for YP* codes
Special Features	

c. Section summary – Qualifiers

Codes beginning 1	GEOGRAPHICAL
Use	To indicate the geographical scope or applicability of book content – such as the location of a travel guide, the setting of a novel, the jurisdiction to which laws apply, <i>etc.</i> (Do not use: to indicate the location or nationality of author, publisher <i>etc.</i>)
Nat Extensions	Typically available for participating countries

Codes beginning 2	LANGUAGE
Use	To indicate the language(s) to which the book content applies – such as a linguistics or literary studies work, phrasebook or dictionary, <i>etc.</i> ; that is, the language a book is about, not necessarily the language it is written in. (Do not use: to indicate the language of the text itself)
Nat Extensions	Not applicable (all codes are shared values)

Codes beginning 3	TIME PERIOD
Use	To indicate the time period range of book content – such as for a history book, memoir or biography, historical fiction, <i>etc.</i> (Do not use: to indicate the year of an annual, road atlas, travel guide, conference proceedings <i>etc.</i>)
Nat Extensions	Typically available for participating countries

Codes beginning 4	EDUCATIONAL PURPOSE
Use	To indicate the curriculum, examination or level for which educational material is specifically designed – such as school textbooks, language learning coursework, revision and study aids, vocational training material. (Do not use: to describe educational institutions a book is about)
Nat Extensions	Typically available for participating countries

Codes beginning 5	INTEREST AGE & SPECIAL INTEREST
Use	To indicate a variety of characteristics relating to content: the particular reading age or level (especially for children’s books); related events (eg religious or public holidays); groups of people that book content is specifically intended for, and/or, in some cases, about (eg for women, for/about religious & ethnic groups, for/about Gays & Lesbians, etc); and to indicate explicit content.
Nat Extensions	Some present

Codes beginning 6	STYLE
Use	To indicate the particular style of artistic or creative expression covered by book content – such as books on artists, art history, music, literary studies, etc. Styles do not include Fiction genres, or styles of writing.
Nat Extensions	Not applicable (all codes are shared values)

4. Worked examples

This section demonstrates the classification of several well-known titles, using the *Thema* scheme.

The Republic, by Plato

The classifications should indicate both the period of philosophy represented and the core philosophical topic.

Assign	QDHA	<i>Ancient philosophy</i>
	QDTS	<i>Social & political philosophy</i>

War and Peace, by Leo Tolstoy

The scope note for FBC instructs it should be used for all fiction written before c 1945. In *Thema*, Geographical and Time Period Qualifiers may be used with fiction where the setting of place and/or time is a significant feature of the work.

Assign	FBC	Classic fiction
	1DTA	Russia
	3MNB	Early 19th century c 1800 to c 1850

Also, if the edition is in any language other than Russian:

	FYT	Fiction in translation
--	-----	------------------------

Harry Potter and the Prisoner of Azkaban, by J.K. Rowling

Fiction written for children should be classified with the YF* codes, and should carry an Interest Age Qualifier wherever possible. (The 'adult edition' should be classified under FM Fantasy.)

Assign	YFH	Children's / Teenage fiction: Fantasy & magical realism
	5AJ	Interest age: from c 8 years

Tintin in America, by Hergé

Thema reflects the fact that, in many markets where graphic novels and comic books are an established and important genre (such as France, Belgium, USA and Japan), the distinction between children's and adult works is less important than the need to group all such items together. Thus, certain graphic novels or comic book series – such as *Tintin*, *Asterix etc* – are the only items specifically intended for children and teenagers that are *not* classified under Y* codes in *Thema*.

Assign	XADC	European comic books: general, classic, all ages
	XQG	Graphic novel / Comic book: action & adventure
	1KBB	United States of America, USA
	5AJ	Interest age: from c 8 years

Lonely Planet: Vienna

It is essential for location-specific travel guides that the most precise available Geographical Qualifier is assigned.

Assign	WTH	<i>Travel & holiday guides</i>
	1DFA-AT-W	<i>Wien / Vienna *</i>

(* for non-users of Austrian National Extensions, this will resolve to 1DFA Austria)

5. Communication of *Thema* Subject Categories and Qualifiers

a. In ONIX

In ONIX 3.0, a book about a particular artistic school of late 19th Century French painting might be classified and communicated like this:

```
<!-- ***** ONIX 3.0 ***** -->
<Subject>
  <MainSubject/>
  <SubjectSchemeIdentifier>93</SubjectSchemeIdentifier> <!-- 93 = Thema subject, see List 27 -->
  <SubjectSchemeVersion>1.0</SubjectSchemeVersion> <!-- version 1.0 -->
  <SubjectCode>AGA</SubjectCode> <!-- AGA = History of art -->
</Subject>
<Subject>
  <SubjectSchemeIdentifier>93</SubjectSchemeIdentifier>
  <SubjectSchemeVersion>1.0</SubjectSchemeVersion>
  <SubjectCode>AFCL</SubjectCode> <!-- AFCL = Paintings and painting in oil -->
</Subject>
<Subject>
  <SubjectSchemeIdentifier>94</SubjectSchemeIdentifier> <!-- 94 = Thema geographical qualifier -->
  <SubjectSchemeVersion>1.0</SubjectSchemeVersion>
  <SubjectCode>1DD-FR-FB</SubjectCode> <!-- 1DD-FR-FB = Finistere -->
  <!-- note 1DD = France if the national extension is unknown -->
</Subject>
<Subject>
  <SubjectSchemeIdentifier>96</SubjectSchemeIdentifier> <!-- 96 = Thema time period qualifier -->
  <SubjectSchemeVersion>1.0</SubjectSchemeVersion>
  <SubjectCode>3MNQX</SubjectCode> <!-- 3MNQX = 1880s -->
</Subject>
<Subject>
  <SubjectSchemeIdentifier>99</SubjectSchemeIdentifier> <!-- 99 = Thema style qualifier -->
  <SubjectSchemeVersion>1.0</SubjectSchemeVersion>
  <SubjectCode>6SV</SubjectCode> <!-- 6SV = Synthetism -->
</Subject>
```

This ONIX 3.0 subject classification might be accompanied by keywords such as 'Pont Aven' or 'Post-impressionist', and 'name as subject' composites for Paul Gauguin and Émile Bernard. This example is taken from the *ONIX 3.0 Implementation and Best Practice Guide*.

In ONIX 2.1, there are only minor differences: the first <Subject> composite would use <MainSubject> instead of <Subject> and <MainSubjectSchemeIdentifier> instead of <SubjectSchemeIdentifier>, and it would omit the <MainSubject/> flag:

```
<!-- ***** ONIX 2.1 ***** -->
<MainSubject>
  <MainSubjectSchemeIdentifier>93</MainSubjectSchemeIdentifier> <!-- 93 = Thema subject, see Lists 26 and 27 -->
  <SubjectSchemeVersion>1.0</SubjectSchemeVersion>           <!-- version 1.0 -->
  <SubjectCode>AGA</SubjectCode>                             <!-- AGA = History of art -->
</MainSubject>
<Subject>
  <SubjectSchemeIdentifier>93</SubjectSchemeIdentifier>
  <SubjectSchemeVersion>1.0</SubjectSchemeVersion>
  <SubjectCode>AFCL</SubjectCode>                           <!-- AFCL = Paintings and painting in oil -->
</Subject>
<Subject>
  <SubjectSchemeIdentifier>94</SubjectSchemeIdentifier> <!-- 94 = Thema geographical qualifier -->
  <SubjectSchemeVersion>1.0</SubjectSchemeVersion>
  <SubjectCode>1DD-FR-FB</SubjectCode>                     <!-- 1DD-FR-FB = Finistere -->
                                                           <!-- note 1DD = France if the national extension is unknown -->
</Subject>
<Subject>
  <SubjectSchemeIdentifier>96</SubjectSchemeIdentifier> <!-- 96 = Thema time period qualifier -->
  <SubjectSchemeVersion>1.0</SubjectSchemeVersion>
  <SubjectCode>3MNQX</SubjectCode>                         <!-- 3MNQX = 1880s -->
</Subject>
<Subject>
  <SubjectSchemeIdentifier>99</SubjectSchemeIdentifier> <!-- 99 = Thema style qualifier -->
  <SubjectSchemeVersion>1.0</SubjectSchemeVersion>
  <SubjectCode>6SV</SubjectCode>                           <!-- 6SV = Synthetism -->
</Subject>
```

b. In formats other than ONIX

Thema code values or headings may of course be transmitted between trading partners in other message formats, displayed on websites, and used in catalogues and in promotional material. Headings may be displayed in any language as appropriate, but to increase intelligibility and interoperability, it is always useful to display the code along with any *Thema* heading.

The syntax of *Thema* codes is as follows:

- Subject codes: one to four upper case letters A–Z, or four upper case letters followed by one or more digits 1–9. The longest codes currently defined are six characters (*eg* QRFB21 – Tibetan Buddhism) though there is no defined limit on the number of digits
- Shared value Qualifiers: one digit 1–6, followed by one or more upper case letters. The longest codes currently defined are seven characters (*eg* 1KJWVVK – Virgin Islands (US)), though there is no defined limit on the number of letters
- National Extension Qualifiers: one digit 1–6, followed by one or more upper case letters, followed by a hyphen, two upper case letters forming an ISO country code, and another hyphen, followed by zero or more upper case letters. The longest code currently defined is thirteen characters (1KBB-US-WPNWS – Seattle) though there is no defined limit on the number of letters (either before or after the country code).

For the avoidance of doubt, codes like 1ABC-DE- and 1ABC-DE-F could plausibly be (German) national extensions, whereas 1ABC-DE (without a final hyphen) is not allowed. In the hierarchy, the 'parent' of 1ABC-DE-F may be either 1ABC-DE- or 1ABC.

For developers wishing to check the plausibility of a *Thema* code without matching against the full list of codes, the following regular expression patterns may be used:

Subject codes ^([A-Z]{1,4}|[A-Z]{4}[1-9]{1,4})\$
Qualifiers (including extensions) ^[1-6][A-Z]{1,7}(-[A-Z]{2}-[A-Z]{0,8})?\$\$

6. Resolution of National Extensions

As stated, all National Extension codes are suffixes, beginning at the first hyphen, that are added to ordinary Shared Value Qualifier codes. If a user encounters or receives a National Extension code that is not known or recognised, it can be interpreted meaningfully by resolving the code string after truncation back to the code value before the first hyphen. This will always be a Shared Value.

Thus if 1KBC-CA-SGA (which indicates 'Central Saskatchewan: Saskatoon') is encountered and not recognised, it can be resolved and understood as 1KBC (Canada).

This resolution process can be done manually, or automatically by the application of simple rules to manipulate incoming data.

Howard Willows
Chair of the Thema ISC

Graham Bell
EDItEUR
October 2013